INTERNATIONAL LAW ASSOCIATION

INTERNATIONAL LAW ASSOCIATION
COMMITTEE ON ARTIFICIAL INTELLIGENCE AND TECHNOLOGY LAW
STATEMENT ON AI GOVERNANCE AND EXISTENTIAL RISK

4 November 2026

Statement

We, the ILA Committee on Artificial Intelligence and Technology Law, express serious concern about the risks arising from increasingly capable AI systems, including the possibility of severe harm across borders and sustained loss of effective human control. We recognise the benefits AI can bring to economic prosperity and social well-being, especially in science, healthcare and public services. We, however, stress that its development and use must serve human well-being, uphold international law and protect the lives, rights and dignity of present and future generations.

We recommend that AI regulation and governance address deliberate misuse and misalignment, where a system pursues goals or behaves in ways that conflict with human intentions and values, as well as structural risks arising from competitive pressures, concentration of capabilities and weakened State capacity. These are distinct problems and require attention to human conduct, system behaviour and the conditions in which AI is developed and used. Existential risk concerns threats to humanity's survival or its lasting ability to determine its future. Catastrophic and existential risks remain subject to significant scientific uncertainty. Uncertainty about such outcomes must be acknowledged, not presented as proof that they are inevitable or as assurance that existing safeguards are sufficient.[1]

We advocate that such uncertainty should not delay proportionate preventive measures where there are credible grounds to anticipate grave or irreversible harm to individuals, societies or future generations. Such measures should be accompanied by efforts to obtain information for a fuller assessment and reviewed within a reasonable period as evidence develops. This approach draws on the precautionary approach, the due-diligence duty to prevent significant transboundary harm and positive human rights obligations to protect life.[2] Recent reported incidents involving AI agents gaining unauthorised access call for States and international bodies to reconsider containment, oversight and incident response throughout the AI lifecycle, including training, evaluation, release and post-release operation.[3] We call for an international, scientifically independent and inclusive process, building on existing mechanisms, to assess and exchange evidence on advanced AI capabilities and associated risks, including their likelihood and severity and the pace of capability development, through the arrangements in Recommendation 4.

We note the UN Secretary-General's call at the General Assembly on 22 September 2026 for shared safety information, cooperation on testing and evaluation, common safeguards, responsible pacing of AI development

---

[1] International AI Safety Report 2026 (3 February 2026), section 2.2, discussion of misalignment and loss-of-control risks.

[2] Rio Declaration on Environment and Development, UN Doc A/CONF.151/26 (Vol I) (1992), annex I, principle 15; cf Agreement on the Application of Sanitary and Phytosanitary Measures, art 5.7, on obtaining further information and reviewing provisional measures. The SPS provision is an analogy, not a rule governing AI. On prevention, the no-harm principle and due diligence, see Pulp Mills on the River Uruguay (Argentina v Uruguay) (Judgment) [2010] ICJ Rep 14, paras 101 and 204–205. Environmental authorities inform these AI recommendations by analogy. On positive obligations to protect life, see Human Rights Committee, General comment No 36, UN Doc CCPR/C/GC/36, paras 7, 18 and 21–22.

[3] Independent International Scientific Panel on AI, AI Agents, Misalignment and the Risk of Losing Human Control (21 September 2026, advance unedited version 1); UNRIC, 'UN panel calls for stronger safeguards as AI agents advance' (21 September 2026). These sources distinguish observed incidents from forecasts of future loss of control.

and a multilateral risk-management framework with independent oversight. We also take note of the Independent International Scientific Panel on Artificial Intelligence's advance assessment of 21 September 2026, which cautions that containing one incident does not establish that increasingly capable systems will remain controllable. These developments demand sustained scientific scrutiny and international legal attention.[4] We further note the September 2026 call by State leaders and the President of the European Commission for mandatory testing before deployment, common standards, shared incident reporting and exploration of an international institution for standards, verification and consultation when capability thresholds are crossed.[5]

We call for a coordinated global response. Separate national and institutional efforts must be connected by shared information, cooperation and clear responsibilities. A common approach need not impose a single legal model, but it must prevent serious risks from falling between jurisdictions. Governance should preserve beneficial research, development and uses of AI, avoid unnecessary barriers and respect different national legal approaches while supporting interoperability for cross-border risks. States with leading AI capacity should act promptly to regulate their AI industries and establish effective rules on accountability and liability for harmful consequences, including unforeseen consequences.

Our work is carried forward through five subcommittees, whose research and internal deliberations inform these recommendations.[6]

Recommendations

On the basis of our work to date, we present the following preliminary recommendations.

1. Require safety evidence. We recommend independent assessment before training, testing, release or deployment that presents a credible risk of grave harm, with reassessment as capabilities and operating conditions change. Requirements should reflect demonstrated and reasonably foreseeable capabilities, rather than depend on whether a system is described as AGI or ASI.[7] Assessments should cover the model, its tools and permissions, its operating environment and interactions among AI agents, including risks of unauthorised access, deception, attempts to evade oversight and failures of control. Developers and operators should provide documented evidence that safeguards are effective, with secure access for competent authorities and independent assessors. They should demonstrate that access restrictions, monitoring and intervention procedures work in practice, including the ability to stop unsafe operation. Where a credible risk of grave harm cannot be adequately controlled, deployment should be withheld, restricted or suspended, subject to proportionate and evidence-based review.[8] Assessments should address loss-of-control risks involving weapons, nuclear command and control and autonomous cyber operations, consistently with applicable

[4] UN Secretary-General, address to the general debate of the eighty-first General Assembly (22 September 2026), SG/SM/23292; Independent International Scientific Panel on AI, AI Agents, Misalignment and the Risk of Losing Human Control: Evidence from the OpenAI-Hugging Face Incident (21 September 2026, advance unedited version 1).

[5] 'A Call for Control of Frontier AI Models' (21 September 2026), calls 1–3, published by the Office of the President of Finland and endorsed by State leaders and the President of the European Commission.

[6] Global Norms, Governance and Comparative Regulation; AI and Human Rights; Accountability and Cross-Border Governance; AI and Sectoral Legal Frameworks; Sustainable and Ethical AI Development.

[7] AGI (artificial general intelligence) refers here to AI with human-level general capabilities across a broad range of cognitive tasks. ASI (artificial superintelligence) refers here to AI exceeding human capabilities across such tasks. These descriptive labels do not establish the assessment threshold.

[8] For existing approaches to evaluation and monitoring, see Regulation (EU) 2024/1689, arts 55, 72 and 73, respectively concerning general-purpose AI models with systemic risk and monitoring and incident reporting for high-risk AI systems; 'Frontier AI Safety Commitments, AI Seoul Summit 2024' (21 May 2024), Outcome 1, commitments I and V (voluntary commitments). Their scope and legal status differ from the Committee's recommendations.

international law, including international humanitarian law, and with regard to the Convention on Certain Conventional Weapons process and UN General Assembly resolutions on autonomous weapons.[9]

Safety governance should include States and communities exposed to AI risks even when they do not host major developers, and should guard against conflicts of interest, regulatory capture and unjustified barriers to independent research and beneficial innovation. Safety measures should not exclude developing countries from secure access to capability-evaluation tools and independent research or entrench existing inequalities.

2. Preserve human control and legal accountability. We call on States, developers and institutions using AI to ensure that human oversight carries the authority, information and practical means to intervene. An AI system's departure from human instructions should not, by itself, absolve relevant persons or institutions of legal responsibility; nor should its compliance with developers' instructions or specifications justify unlawful harm. Deploying institutions should maintain oversight mechanisms and the technical and operational capacity to stop activities presenting a credible risk of grave harm. Legal responsibility must remain with the relevant persons and institutions, with accessible procedures to challenge decisions and seek remedies.[10] Contractual allocations of risk among developers, deployers and users should not leave affected third parties without an effective remedy. States should require the preservation of records identifying accountable personnel and including evaluation results and deployment monitoring logs sufficient to investigate harmful conduct and determine liability, subject to privacy and security safeguards.

3. Provide proportionate and risk-based regulatory safeguards. We call on States to provide legally defined powers for competent authorities to withhold, restrict or suspend relevant AI development, training, testing, release or deployment, or require withdrawal of a released system, where a credible risk of grave harm cannot be adequately controlled. Measures should specify their scope, reasons and conditions for resumption, and be subject to independent scrutiny and timely review. Resumption should depend on independently verified compliance with published safety requirements. States should assess the need for a moratorium, ban or other appropriate measures concerning certain uses of AI systems that they consider incompatible with human rights, democracy or the rule of law.[11] Where narrower measures cannot adequately address serious cross-border risks, States should coordinate broader temporary restrictions. This recommendation extends to pre-release research and development and to defence and military systems, subject to applicable international law.[12] Restrictions should depend on assessed risk rather than developers' nationality or location, and should not constitute arbitrary or unjustifiable discrimination or disguised restrictions on trade or scientific cooperation. Secure remediation and safety research should remain possible under appropriate controls, including access for developing countries to capability-evaluation tools and independent research.

[9] See CCW, Final report of the 2019 Meeting of the High Contracting Parties, UN Doc CCW/MSP/2019/9, annex III, guiding principles (a), (b), (d) and (g); UNGA Res 79/62, 'Lethal autonomous weapons systems' (2 December 2024).

[10] UNESCO, Recommendation on the Ethics of Artificial Intelligence (23 November 2021), paras 35–36 and 42–43.

[11] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No 225 (5 September 2024), art 16(4). The provision requires assessment of the need for measures concerning certain incompatible uses; the Committee's broader recommendations on regulatory powers are proposals in this Statement.

[12] Council of Europe Framework Convention, arts 3(3)–(4): specified research and development on systems not yet available for use is excluded, subject to the exception for testing or similar activities capable of interfering with human rights, democracy or the rule of law; national defence is excluded. The Committee recommends broader coverage of these activities where a credible risk of grave harm arises. This addresses gaps in the coverage of pre-release training and evaluation.

4. Establish cross-border arrangements for prevention and response. As the incidence and impact of AI harm traverse jurisdictional boundaries, we urge States and international organisations to establish interoperable procedures for mandatory reporting of serious AI incidents, significant near misses and significant unexpected capabilities or behaviours discovered after release, preservation of evidence and immediate notification to affected operators and competent authorities in affected States. Initial notification should be made when there are reasonable grounds to suspect a serious incident, without awaiting completion of an investigation or confirmation of cause or attribution. Designated contact points should support early warnings, joint investigation and coordinated response, with lawful pathways for transferring incident-related data and evidence across borders while protecting privacy and legitimate security interests. We also call for clarity on jurisdiction, responsibility and access to remedy where harm crosses borders. Cooperation should connect existing institutions and scientific bodies rather than create parallel processes with no means of working together. These arrangements should support prompt, independent assessment of serious incidents, observed capabilities, operating conditions and the adequacy of safeguards. International scientific bodies should support common assessments and evaluation standards, while enforcement decisions should rest with authorities possessing the necessary legal mandate. States should act within their existing powers while developing further national and international arrangements.

In developing these arrangements, States should distinguish the responsibility of private persons and institutions from the international responsibility of States. A developer's nationality or location does not, by itself, establish attribution of its conduct to a State. States should exercise due diligence in regulating private actors within their jurisdiction or control whose activities risk significant transboundary harm. Lack of attribution of private conduct does not exclude responsibility for a State's own failure to discharge applicable duties of prevention and cooperation.[13] Further work should clarify these duties, jurisdiction, access to evidence and effective remedies, while distinguishing existing legal obligations from proposals for the development of international law.

5. Protect livelihoods, rights and the environment. We stress that preventing existential harm does not excuse neglect of discrimination, privacy violations, unsafe working conditions or environmental damage. States should ensure that laws permitting AI practices comply with their obligations to protect fundamental rights and human dignity, which cannot be displaced by convenience, income or formal legalisation alone. We call for assessments of significant effects on human rights, including labour rights, and the environment, proportionate to the nature, scale and reasonably foreseeable risks of the application, with particular attention to vulnerable groups, children in particular, and individuals in vulnerable situations more broadly. Workers and affected communities should participate in decisions about deployment and have access to sufficient training, social protection and remedies. States should also provide training for public officials on AI risks and rights. Those least able to influence deployment must not be left to bear its costs.[14]

6. Ensure meaningful global participation and respect Indigenous Peoples' rights. We call for participation across regions, genders and racial and ethnic groups, with practical support for developing countries and underrepresented communities. We urge respect for Indigenous Peoples' data sovereignty and their authority over the use of their knowledge, languages and community data, alongside their rights recognised under applicable international and domestic law, including rights concerning traditional knowledge and culturally significant data, and with regard to continuing developments in Indigenous data governance. Participation should take place through their own representative institutions, with free, prior and informed consent where applicable. Inclusion must extend to setting priorities, assessing risks and sharing benefits, not merely attendance at international discussions.[15]

---

[13] See, by analogy, Obligations of States in respect of Climate Change (Advisory Opinion, 23 July 2025), paras 280–283, 293–294 and 428, distinguishing a State's own regulatory omissions from attribution of private conduct; ITLOS, Request for an Advisory Opinion submitted by the Commission of Small Island States on Climate Change and International Law (21 May 2024), Case No 31, paras 239–243. These environmental and climate authorities inform the recommended approach to AI; they do not decide AI-specific obligations.

[14] ILO, Global case studies of social dialogue on AI and algorithmic management, Working Paper 144 (10 July 2025).

[15] United Nations Declaration on the Rights of Indigenous Peoples, UNGA Res 61/295 (13 September 2007), arts 18, 19 and 31.

Call for cooperation

Building on the research and internal deliberations of its subcommittees, the Committee intends first to develop a definition of 'credible risk of grave harm', followed by practical guidance on risk thresholds, conditions for suspension and resumption, post-release monitoring and re-evaluation, cross-border incident reporting and accountability and remedies. This work will distinguish existing international obligations from proposals for further development and will be submitted for Committee consideration and endorsement.

We invite governments, international organisations, the legal and scientific communities, industry and civil society to contribute to this work. We strongly believe that AI must serve humanity. International cooperation must protect human life, preserve human agency and ensure that the benefits of AI are shared without sacrificing the rights or safety of others.

About the Committee

Established in May 2025, the International Law Association (ILA) Committee on Artificial Intelligence and Technology Law is a global committee comprising 44 members from 30 countries and jurisdictions across six continents. Its membership includes legal scholars, practitioners and judicial officers from different legal traditions. The Committee is chaired by Dr Jon Truby, holder of the UNESCO Chair on AI Law & Sustainability at the Centre for International Law, National University of Singapore. Its programme combines comparative legal research with international dialogues and conferences, including its recent sessions at the 2026 ILA Biennial Conference in Vienna. The Committee forms part of the ILA, founded in 1873, which holds consultative status with a number of United Nations specialised agencies.[16]

Members participate in the Committee in their individual expert capacities. The views expressed in this Statement do not necessarily reflect the positions of their governments, employers or institutions.

To learn more about the Committee's work or to contribute, please contact Sheikh Sultan Aadil Huque (Secretary): aadil.huque@nus.edu.sg

[16] Committee Secretariat, Vienna event report, August 2026; ILA, About Us and Committee mandate.